

SISTEM SCRAPING DAN KLASIFIKASI DATA PERCAKARAPAN SAHAM PADA APLIKASI TELEGRAM DENGAN ALGORITMA SUPPORT VECTOR MACHINE

SCRAPING AND CLASSIFICATION SYSTEM OF STOCK CONVERSATION DATA IN TELEGRAM APPLICATION WITH SUPPORT VECTOR MACHINE ALGORITHM

Al Agias Bayu Asa¹, Burhanuddin Dirgantoro², Casi Setianingsih³

^{1,2,3} Universitas Telkom, Bandung

aloysiusagias@student.telkomuniversity.ac.id¹, burhanuddin@telkomuniversity.ac.id²,

setiacasie@telkomuniversity.ac.id³

Abstrak

Pemilik saham dapat dikatakan sebagai pemilik perusahaan, semakin besar saham yang dimiliki maka semakin besar pula kekuasaannya di perusahaan tersebut, sehingga tidak sedikit orang yang mulai belajar bagaimana membeli saham. Saham sangat terpengaruh oleh keadaan yang ada di dunia ini, maka dari itu seseorang yang mempunyai saham harus mengetahui kabar terbaru yang dapat dibaca di berita dan kabar burung. Sehingga penulis bertujuan untuk dapat menganalisis suatu pesan sehingga dapat menghasilkan klasifikasi dari pesan tersebut untuk selanjutnya dapat digunakan untuk tolok ukur dalam membeli saham. Penelitian ini dilakukan untuk membuat sistem yang akan mengklasifikasikan pesan pada aplikasi pesan Telegram. Dengan menggunakan algoritma *Support Vector Machine*, sistem dapat mengklasifikasikan pesan yang dikirim di Telegram. Karena Telegram merupakan aplikasi yang terdapat berbagai komunitas di dalamnya, dan salah satu nya adalah komunitas saham. Sistem pada penelitian Tugas Akhir telah di uji menggunakan partisi data uji 95% dan data latih 5%, dan juga parameter $C=1$ dan $\text{Gamma}=1$ mendapatkan akurasi sebesar 90%, *precision* sebesar 93%, *recall* sebesar 89%, dan juga *F1-Score* sebesar 91%.

Kata Kunci: klasifikasi, analisis, saham, SVM.

Abstract

Shareholders can be said to be the owners of the company, the larger the stocks owned, the greater the power in the company, so that not a few people are starting to learn how to buy stocks. Stocks are greatly affected by the conditions that exist in this world, therefore someone who owns stocks must know the latest news that can be read in the news and rumors. So the author aims to be able to analyze a message so that it can produce a classification of the message which can then be used as a guidelines in buying stocks. This research was conducted to create a system that will classify messages on the Telegram messaging application. By using the Support Vector Machine algorithm, the system can classify messages sent on Telegram. Because Telegram is an application that has various communities in it, and one of them is the stock community. The system in the Final Project has been tested using a 95% test data partition and 5% training data, and also the parameters $C=1$ and $\text{Gamma}=1$ get an accuracy of 90%, precision of 93%, recall of 89%, and also F1- Score of 91%.

Keywords: classification, analysis, stock, SVM.

1. Pendahuluan

Investor perlu mendapatkan keuntungan dari berinvestasi, keuntungan yang didapatkan dari membeli saham yaitu *capital gain* dan dividen [1]. Jenis keuntungan Capital Gain memerlukan prediksi yang tepat dan informasi yang banyak untuk mendapatkan informasi mengenai perusahaan mana yang tengah hangat dibicarakan, karena secara tidak langsung berita juga berpengaruh ke naik dan turunnya harga saham. Atas dasar itulah para investor saham harus banyak membaca berita untuk menggali informasi dengan perkembangan di dunia ini. Investor harus cepat tanggap dalam berbagai perubahan, faktor yang dapat mempengaruhi tentang tingkat perolehan saham antara lain faktor mikro dan makro ekonomi [2]. Mendapatkan informasi mengenai faktor mikro dan makro ekonomi salah satunya adalah bertukar informasi mengenai isu dunia dan perusahaan di sebuah platform sehingga para investor atau *trader* saham dapat saling berinteraksi satu sama lain. Salah satu platform yang memungkinkan ada nya pertukaran informasi adalah Telegram, karena komunitas di Telegram mudah dicari hanya dengan menggunakan kata kunci lalu dapat masuk kedalam komunitas dengan amat mudah. Atas dasar hal tersebut maka penulis menargetkan platform grup Telegram sebagai target penelitian

2. Dasar Teori

2.1 Saham

Saham Merupakan bukti bahwa seseorang telah menanamkan atau ikut serta memberikan modal pada suatu perusahaan, dengan mempunyai saham maka investor dapat memiliki keuntungan sesuai dengan saham yang

ditanamnya. Apabila suatu saham mempunyai kelebihan peminat, maka harga nya akan cenderung naik, sedangkan apabila kelebihan penawaran maka harga nya pun akan cenderung turun [3].

2.2 Web Scraping

Web Scraping merupakan suatu proses yang dilakukan untuk mengambil data yang ada pada sebuah dokumen di internet, dokumen tersebut biasa nya berbentuk HTM [4]. Data Scraping merupakan istilah yang digunakan untuk mendeskripsikan pengambilan data dari file elektronik menggunakan program komputer. Web Scraping mendeskripsikan penggunaan program untuk mengambil data dari file HTML di internet [5].

2.3 Support Vector Machine

Algoritma SVM ini pertama kali diperkenalkan pada tahun 1992 oleh Vapnik, SVM ini merupakan metode yang ada pada Machine learning yang bertujuan untuk memisahkan antara dua buah klasifikasi pada data [6]. Dua buah klasifikasi dipisah menggunakan sebuah hyperplane, algoritma akan mencoba mencari hyperplane mana yang paling efektif dalam pengelompokan data tersebut sehingga data tersebut dapat dikelompokkan dengan baik.

$$\vec{w} \cdot \vec{x} + b = 0 \quad (2.1)$$

Dengan memperhatikan maksimal nilai jarak hyperplane dengan titik terdekatnya, yaitu $1/\|\vec{w}\|$. Hal tersebut dapat juga dirumuskan sebagai Quadratic Problem atau yang biasa disebut dengan QP Problem, yang berarti mencari suatu titik minimal persamaan(2.4), dan juga dengan memperhatikan constraint persamaan(2.5).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (2.2)$$

$$y_i((\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \forall i \quad (2.3)$$

Masalah tersebut dapat dicari solusi dengan menggunakan berbagai Teknik komputasi, antara lain Lagrange Multiplier.

2.4 POS Tagging

Part-Of-Speech (POS) Tagger merupakan usaha untuk melabeli sebuah kata yang ada didalam teks sesuai dengan kelas kata dan juga fitur morfologinya. Hasil dari upaya tersebut merupakan informasi mengenai definisi dan konteks kata, sehingga proses ini sangat penting dalam proses awal pengolahan kata untuk NLP [7]. Data POS Tagset terdiri dari 23 POS tags. Berdasarkan fakta bahwa kata kerja dan kata benda biasanya lebih penting dari pada kata sifat dan kata keterangan, yang mana kata sifat dan keterangan itu sendiri lebih penting dari pada yang lain seperti kata ganti, preposisi, konjungsi dan lainnya. Maka dari itu fungsi pembobotan akan menjadi seperti berikut :

$$w_{pos}(t) = \begin{cases} w_1 & \text{jika } t \text{ adalah kata kerja/benda} \\ w_2 & \text{jika } t \text{ adalah kata keterangan/sifat} \\ w_3 & \text{lainnya} \end{cases} \quad (2.8)$$

Dimana nilai $w_1 > w_2 > w_3 > 0$. Dan disini penulis menetapkan nilai untuk $w_1 = 5$, $w_2 = 3$, $w_3 = 1$ [8].

2.5 TF-IDF

Pembobotan diperlukan dalam Natural Language Processing, karena dengan pembobotan suatu kata dapat mendapatkan nilai. Salah satu metode pembobotan adalah TF-IDF, metode ini digunakan untuk menghitung atau menentukan bobot daripada setiap kata yang ada paling sering digunakan pada sebuah dokumen. Dengan adanya metode ini, akan dihitung nilai Term Frequency dan juga Inverse Document Frequency [9]. Dalam penghitungan nilai TF-IDF, yang perlu dihitung terlebih dahulu adalah nilai dari TF, setelah dihitung nilai dari TF maka selanjutnya dapat dihitung nilai IDF [10]. TF-IDF karena telah terjadi POS Tagging akan menjadi hamper sama seperti TF-IDF versi asli nya.

$$\text{Weight}_{pos}(t, d) = \begin{cases} \frac{c(t, d) * w_{pos}(t)}{\sum_i c(t, d) * w_{pos}(t_i)} * \left(1 + \log \left(\frac{N}{k}\right)\right) & \text{jika } c(t, d) \geq 1 \\ 0 & \text{lainnya} \end{cases} \quad (2.9)$$

3. Pembahasan

3.1. Gambaran umum Sistem

Algoritma yang akan digunakan adalah algoritma Support Vector Machine yang akan diimplementasikan pada klasifikasi pesan, teks pesan akan diolah menggunakan algoritma tersebut sehingga akan menghasilkan keluaran positif, negatif, dan netral dari berita tersebut.

3.2 Perancangan Sistem

Sistem dirancang dengan menggunakan berbagai varian proses seperti proses pengambilan data, preprocessing data, ekstraksi ciri, pembobotan data, Normalisasi, hingga klasifikasi menggunakan Support Vector Machine.

3.3 Data Retrieval

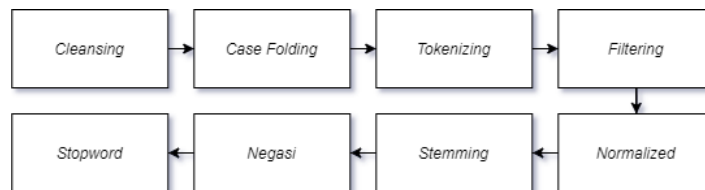
Dataset diambil melalui scraping menggunakan selenium, dalam proses pengambilan dataset tidak menggunakan Telegram BOT Api karena BOT perlu ditambahkan ke grup oleh admin dan tidak bisa langsung masuk tanpa persetujuan admin sehingga pengguna tidak bisa mengklasifikasikan dengan mudah apabila menggunakan bot untuk mengambil dataset yang diperlukan. Data yang diambil dari sebuah pesan adalah nama pengirim, teks pesan, balasan pesan apabila ada dan pengirim balasan. Berikut merupakan contoh dataset yang diambil dari grup telegram yang bernama "The Traders Saham" :

Tabel 3.1. Contoh Data Pesan Telegram

Label	Pesan	Inisial
Positif	Sepertinya EXCL mau mulai terbang..... Guys..... YU sudah akum seminggu..... beli terus...	1
Negatif	bmri kemungkinan bid palsu dr tadi ga kuat, malah turun	-1
Negatif	Sudah paham sebenarnya resiko loss di saham itu ada. Tapi pada saat itu terjadi susah sekali untuk mengambil keputusan cut loss...	-1

3.4 Preprocessing

Setelah mendapatkan data berbentuk teks, maka teks tersebut perlu diolah terlebih dahulu menggunakan library NLTK, pemrosesan awal data dibagi menjadi beberapa bagian. Dapat di perhatikan gambar berikut.



Gambar 3.1. Preprocessing Data

Untuk menggambarkan hasil dari preprocessing maka akan ditampilkan hasil dari data pada tabel di bagian data retrieval, maka hasil nya akan menjadi seperti berikut:

Tabel 3.2. Hasil dari preprocessing

Output
'terbang', 'ayo', 'akumulasi', 'beli'
'bid', 'palsu', 'tidak kuat', 'turun'
'resiko', 'loss', 'susah', 'putus', 'cut', 'loss'

3.5 POS Tagging

POS Tagging dilakukan setelah teks telah di lakukan pre-processing, POS Tagging dilakukan agar tiap kata di kalimat memiliki makna yang sebenarnya dan akan membantu dalam proses pembobotan atau TF-IDF di Langkah berikut nya. Sebagai contoh dataset yang telah memasuki proses preprocessing tersebut kemudian akan diolah dengan POS Tagging untuk mengetahui makna dari tiap kata pada kalimat tersebut.

Tabel 3.3. Hasil dari POS Tagging

Input	Output
'terbang', 'ayo', 'akumulasi', 'beli'	'terbang/VB', 'ayo/NN', 'akumulasi/NN', 'beli/VB'
'bid', 'palsu', 'tidak kuat', 'turun'	'bid/NN', 'palsu/JJ', 'tidak kuat/NN', 'turun/VB'
'resiko', 'loss', 'susah', 'putus', 'cut', 'loss'	'resiko/NN', 'loss/FW', 'susah/JJ', 'putus/NN', 'cut/NN', 'loss/FW'

Setelah mendapatkan hasil dari POS Tagging maka selanjutnya adalah membuat bag of word atau dihitung tiap kata yang muncul akan tetapi dikalikan menggunakan beban pada tiap arti kata seperti syarat yang telah dicantumkan. sebelumnya perlu diperhatikan bahwa kata yang diproses adalah kata yang masuk dalam kamus kata unigram yang sebelumnya telah dibuat, menjadi seperti berikut:

Tabel 3.4. Hasil dari Normalisasi Data

Term	D1	D2	D3
Terbang	5	0	0
Ayo	5	0	0
Akumulasi	5	0	0
Beli	5	0	0
Bid	0	5	0
palsu	0	3	0
Tidak_kuat	0	5	0
Turun	0	5	0
Resiko	0	0	5
Loss	0	0	2
Susah	0	0	3
Putus	0	0	5
Cut	0	0	1

3.6 TF-IDF

Hasil dari *bag of word* tersebut kemudian akan dilakukan perhitungan menggunakan TF-IDF, hasil dari TF-IDF tersebut kemudian akan digunakan untuk perhitungan pada algoritma. Hasil dari perhitungan TF-IDF dapat dilihat pada tabel berikut ini :

Tabel 3.5. Hasil dari TF-IDF

Term	TF_{pos}			Df	IDF	$tf_{td} * idf$		
	D1	D2	D3			D1	D2	D3
Terbang	0.25	0	0	1	1.4771	0.3693	0	0
Ayo	0.25	0	0	1	1.4771	0.3693	0	0
Akumulasi	0.25	0	0	1	1.4771	0.3693	0	0
Beli	0.25	0	0	1	1.4771	0.3693	0	0
Bid	0	0.2777	0	1	1.4771	0	0.4102	0
palsu	0	0.1666	0	1	1.4771	0	0.2461	0
Tidak_kuat	0	0.2777	0	1	1.4771	0	0.4102	0
Turun	0	0.2777	0	1	1.4771	0	0.4102	0
Resiko	0	0	0.3125	1	1.4771	0	0	0.4616
Loss	0	0	0.125	2	1.1761	0	0	0.147
Susah	0	0	0.1875	1	1.4771	0	0	0.2769
Putus	0	0	0.3125	1	1.4771	0	0	0.4616
Cut	0	0	0.0625	1	1.4771	0	0	0.0923

Setelah di lakukan TF-IDF maka berikutnya akan dilakukan perhitungan dengan normalisasi, tujuan dari perhitungan ini adalah agar nilai dari TF-IDF akan lebih tertata rapi, hasil dari normlaisasi data dapat dilihat pada tabel berikut :

Tabel 3.6. Hasil dari Normalisasi

Term	w_{ij}		
	D1	D2	D3
Terbang	0.6770	0	0
Ayo	0.6770	0	0
Akumulasi	0.6770	0	0
Beli	0.6770	0	0
Bid	0	0.6729	0
palsu	0	0.4037	0
Tidak_kuat	0	0.6729	0
Turun	0	0.6729	0
Resiko	0	0	0.6322
Loss	0	0	0.2013
Susah	0	0	0.3793
Putus	0	0	0.6322

Term	w_{ij}		
	D1	D2	D3
Cut	0	0	0.1264

3.7 Support Vector Machine

Dalam proses ini akan dilakukan perhitungan untuk mencari nilai dari a dan b yang akan berguna dalam pencarian *hyperplane* terbaik. Langkah pertama yang dilakukan adalah pencarian nilai x menggunakan data dari normalisasi, hasil dari perhitungan dapat dilihat pada tabel berikut :

Tabel 3.7. Hasil dari $x - x_i$

X1-X1	X1-X2	X1-X3	X2-X1	X2-X2	X2-X3	X3-X1	X3-X2	X3-X3
0	0,677	0,677	-0,677	0	0	-0,677	0	0
0	0,677	0,677	-0,677	0	0	-0,677	0	0
0	0,677	0,677	-0,677	0	0	-0,677	0	0
0	0,677	0,677	-0,677	0	0	-0,677	0	0
0	-0,6729	0	0,6729	0	0,6729	0	-0,6729	0
0	-0,4037	0	0,4037	0	0,4037	0	-0,4037	0
0	-0,6729	0	0,6729	0	0,6729	0	-0,6729	0
0	-0,6729	0	0,6729	0	0,6729	0	-0,6729	0
0	0	-0,6322	0	0	-0,6322	0,6322	0,6322	0
0	0	-0,2013	0	0	-0,2013	0,2013	0,2013	0
0	0	-0,3793	0	0	-0,3793	0,3793	0,3793	0
0	0	-0,6322	0	0	-0,6322	0,6322	0,6322	0
0	0	-0,1264	0	0	-0,1264	0,1264	0,1264	0

Hasil dari pencarian nilai x tersebut berikut nya akan digunakan untuk pencarian nilai dari panjang vector, hasil pencarian panjang vector adalah sebagai berikut :

Tabel 3.8. Panjang Vector

$\ x - x_i\ $	$\sqrt{x^2 + y^2 + z^2}$	Hasil
$\ x_1 - x_1\ $	$\sqrt{0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2}$	0
$\ x_1 - x_2\ $	$\sqrt{0,677^2 + 0,677^2 + 0,677^2 + 0,677^2 - 0,6729^2 - 0,4037^2 - 0,6729^2 - 0,6729^2 - 0^2 + 0^2 + 0^2 + 0^2}$	1.831577
$\ x_1 - x_3\ $	$\sqrt{0,677^2 + 0,677^2 + 0,677^2 + 0,677^2 + 0^2 + 0^2 + 0^2 + 0^2 - 0,6322^2 - 0,2013^2 - 0,3793^2 - 0,6322^2 - 0,1264^2}$	1.683163

Berikutnya, data akan digunakan untuk perhitungan berikutnya. Dengan dataset berdimensi 1xN maka dapat didapatkan dimensi baru yaitu NxN, yang mana N adalah jumlah data. Dalam perhitungan ini akan menggunakan nilai $y=0.5$.

$$\begin{aligned}
 K(1,1) &= \exp(-y\|x - x_i\|^2), y = 0,5 \\
 &= \exp(-0.5(0)^2) \\
 &= \exp 0 \\
 &= 1 \\
 K(1,3) &= \exp(-y\|x - x_i\|^2), y = 0,5 \\
 &= \exp(-0.5(1.683163)^2) \\
 &= \exp -1.4165 \\
 &= 0.2426
 \end{aligned}$$

$$\begin{aligned}
 K(1,2) &= \exp(-y\|x - x_i\|^2), y = 0,5 \\
 &= \exp(-0.5(1.831577)^2) \\
 &= \exp -1.6773 \\
 &= 0.1869
 \end{aligned}$$

Setelah dilakukan perhitungan maka didapatkan nilai sebagai berikut :

$$K = \begin{bmatrix} 1 & 0.1869 & 0.2426 \\ 0.1869 & 1 & 0.2835 \\ 0.2426 & 0.2835 & 1 \end{bmatrix}$$

Setelah menghitung kernel RBF, maka berikutnya akan dilakukan pencarian dari nilai y, nilai y merupakan label dari data yang telah di berikan. Kemudian akan dilakukan perhitungan dan akan mendapatkan nilai dari $y_i y_j$ sebanyak N data.

Tabel 3.10. Perhitungan Nilai $y_i y_j$

$$y_i y_j = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

Setelah semua nilai di dapatkan maka berikutnya akan digunakan dalam perhitungan menggunakan rumus lagrange.

$$\begin{aligned}
 \min Ld &= \frac{1}{2} \sum_{i,j} a_i a_j \begin{bmatrix} 1 & 0.1869 & 0.2426 \\ 0.1869 & 1 & 0.2835 \\ 0.2426 & 0.2835 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} - a_1 + a_2 + a_3 \\
 &= \frac{1}{2} \sum_{i,j} a_i a_j \begin{bmatrix} -0.5705 & 0.5705 & 0.5705 \\ -1.1686 & 1.1686 & 1.1686 \\ -1.0409 & 1.0409 & 1.0409 \end{bmatrix} - a_1 + a_2 + a_3 \\
 &= \frac{1}{2} \left(\begin{aligned} &-0.5705a_1^2 - 1.1686a_2a_1 - 1.0409a_3a_1 + 0.5705a_1a_2 \\ &+ 1.1686a_2^2 + 1.0409a_3a_2 + 0.5705a_1a_3 \\ &+ 1.1686a_2a_3 + 1.0409a_3^2 \end{aligned} \right) - a_1 + a_2 + a_3 \\
 &= \frac{1}{2} \left(\begin{aligned} &-0.5705a_1^2 - 0.5981a_1a_2 - 0.4704a_1a_3 + 1.1686a_2^2 \\ &+ 2.2095a_2a_3 + 1.0409a_3^2 \end{aligned} \right) - a_1 + a_2 + a_3 \\
 &= \left(\begin{aligned} &-0.2853a_1^2 - 0.2991a_1a_2 - 0.2352a_1a_3 + 0.5843a_2^2 \\ &+ 1.1048a_2a_3 + 0.5205a_3^2 \end{aligned} \right) - a_1 + a_2 + a_3
 \end{aligned}$$

Hasil dari perhitungan tersebut kemudian akan digunakan dalam perhitungan turunan untuk mendapatkan nilai minimum dari lagrange.

1. $\frac{\partial L}{\partial a_1} = -0.2853a_1 - 0.2991a_2 - 0.2352a_3$
2. $\frac{\partial L}{\partial a_2} = -0.2991a_1 + 0.5843a_2 + 1.1048a_3$
3. $\frac{\partial L}{\partial a_3} = -0.2352a_1 + 1.1048a_2 + 0.5205a_3$
4. $a_1 = a_2 + a_3$

Hasil dari turunan akan menghasilkan persamaan, dari persamaan tersebut dapat dilakukan pencarian untuk tiap nilai a.

$$\begin{aligned}
 1-4. \frac{\partial L}{\partial a_1} &= -0.2853a_2 - 0.2853a_3 - 0.2991a_2 - 0.2352a_3 \\
 &= -0.5844a_2 - 0.5205a_3 = 0 \\
 2-4. \frac{\partial L}{\partial a_2} &= -0.2991a_2 - 0.2991a_3 + 0.5843a_2 + 1.1048a_3 \\
 &= 0.2852a_2 + 0.8057a_3 = 0 \\
 3-4. \frac{\partial L}{\partial a_3} &= -0.2352a_2 - 0.2352a_3 + 1.1048a_2 + 0.5205a_3 \\
 &= 0.8696a_2 + 0.2853a_3 = 0
 \end{aligned}$$

Dari persamaan didapatkan hasil akhir adalah $a_1 = 1$, $a_2 = 1$, and $a_3 = 0$. Di karenakan $a_3 = 0$ maka hyperplane dapat di cari menggunakan a_1 dan a_2 .

4. Implementasi dan Pengujian Sistem

4.1 Pengujian Partisi Data

Dalam pengujian ini dilakukan pembagian data sesuai dengan yang telah ditentukan, total data yang digunakan adalah 1847 data.

Tabel 4.1. Hasil Pengujian Partisi Data

Pengujian	Training	Testing	Hasil			
			Accuracy	Precision	Recall	F1-score
1	50%	50%	0.81	0.83	0.75	0.78
2	55%	45%	0.80	0.82	0.74	0.77
3	60%	40%	0.81	0.84	0.75	0.78
4	65%	35%	0.82	0.84	0.78	0.80
5	70%	30%	0.82	0.84	0.77	0.80
6	75%	25%	0.82	0.85	0.79	0.82
7	80%	20%	0.84	0.86	0.79	0.82
8	85%	15%	0.85	0.87	0.81	0.83
9	90%	10%	0.87	0.91	0.85	0.87
10	95%	5%	0.90	0.93	0.89	0.91

Dapat dilihat bahwa semakin kecil data uji maka semakin besar juga nilai akurasi yang didapatkan, hasil tertinggi kemudian akan digunakan dalam pengujian berikut nya. Disini hasil pengujian tertinggi adalah pembagian partisi data latih 95% dan data uji 5% dengan akurasi 90%.

4.2 Pengujian Parameter C

Dalam pengujian ini akan dilakukan percobaan menggunakan nilai C dari 0.01 sampai 10.000.000, dengan menggunakan partisi data dari pengujian sebelumnya maka didapatkan hasil sebagai berikut:

Tabel 4.2. Hasil Pengujian Parameter C

Pengujian	C	Hasil			
		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
1	0.01	0.47	0.16	0.33	0.21
2	0.1	0.59	0.81	0.47	0.46
3	1	0.90	0.93	0.89	0.91
4	10	0.90	0.93	0.89	0.90
5	100	0.88	0.91	0.86	0.88
6	1000	0.88	0.91	0.86	0.88
7	10000	0.88	0.91	0.86	0.88
8	100000	0.88	0.91	0.86	0.88
9	1000000	0.88	0.91	0.86	0.88
10	10000000	0.88	0.91	0.86	0.88

Dapat dilihat dari data diatas terdapat dua nilai terbaik yaitu C = 1 dan C = 10, akan tetapi C = 1 memiliki nilai F1-Score lebih tinggi walaupun hanya sedikit. Maka akan digunakan nilai C = 1 untuk pengujian berikutnya.

4.4 Pengujian Parameter Gamma

Pengujian ini dilakukan dengan menggunakan rentang Gamma antara lain 0.001 sampai 10, dengan menggunakan partisi data dan nilai parameter C dari percobaan sebelumnya.

Tabel 4.3. Hasil Pengujian Parameter Gamma

Pengujian	Gamma	Hasil			
		<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
1	0.0001	0.47	0.16	0.33	0.21
2	0.001	0.47	0.16	0.33	0.21
3	0.01	0.47	0.16	0.33	0.21
4	0.1	0.80	0.90	0.76	0.79
5	1	0.90	0.93	0.89	0.91
6	10	0.75	0.87	0.69	0.73

Dapat dilihat pada grafik bahwa nilai gamma terbaik yang dilakukan pada percobaan adalah gamma = 1 dengan parameter C=1 dan partisi data 95% untuk data latih dan 5% untuk data uji dengan nilai akurasi 90%.

5. Kesimpulan

Kesimpulan yang dapat ditarik dalam percobaan ini adalah sebagai berikut :

1. Sistem scraping dan klasifikasi data percakapan saham pada aplikasi Telegram dengan algoritma Support Vector Machine telah berhasil mengklasifikasikan pesan positif, negatif, dan juga netral.
2. Support Vector Machine yang diimplementasikan berhasil melakukan klasifikasi dari pesan grup telegram dengan performa terbaik yang didapatkan dari hasil pengujian adalah dengan perbandingan data latih sebesar 95% dan data uji sebesar 5% dan parameter C dan Gamma terbaik dalam penelitian ini adalah C = 1 dan juga Gamma = 1 dengan akurasi sebesar 90%, precision sebesar 93%, recall sebesar 89%, dan juga F1-Score sebesar 91%.

Referensi:

- [1] M. Nurhayati, ""Profitabilitas, likuiditas dan ukuran perusahaan pengaruhnya terhadap kebijakan dividen dan nilai perusahaan sektor non jasa," *Jurnal Keuangan & Bisnis Program Studi Magister Manajemen Sekolah Tinggi Ilmu Ekonomi Harapan*, vol. 5, pp. 144 - 153, 2013.
- [2] S. L. Abdes, Z. Puspitaningtyas and A. Prakoso, "Pengaruh Inflasi, Kurs Rupiah dan Suku Bunga terhadap Return Saham Perusahaan Manufaktur," *Jurnal Akuntansi Keuangan dan Bisnis*, vol. 10, no. 2, pp. 8-16, 2017.
- [3] I. Novitasari, D. Budiadi and A. D. Limatara, "Analisis Stock Split Terhadap Harga Saham PT Jaya Real Property Tahun 2010-2016," *CAHAYA AKTIVA*, vol. 10, no. 1, pp. 9-18, 2020.
- [4] J. Ahmat, L. A. Abfillah and Suryayusra, "Penerapan teknik web scraping pada mesin pencari artikel ilmiah.," 2014.

- [5] N. R. Haddaway, "The use of web-scraping software in searching for grey literature.," *Grey J*, vol. 11, no. 3, 2015.
- [6] A. S. Nugroho, A. B. Witarto and D. Handoko, "Support vector machine," *Kuliah Umum IlmuKomputer.Com*, 2003.
- [7] M. Kamayani, "Perkembangan Part-of-Speech Tagger Bahasa Indonesia," *Jurnal Linguistik Komputasional*, vol. 2, no. 2, 2019.
- [8] R. Xu, "POS Weighted TF-IDF Algorithm and its Application for an MOOC Search Engine," *International Conference on Audio, Language and Image Processing*, 2014.
- [9] A. A. Maarif, "Penerapan Algoritma TF-IDF untuk Pencarian Karya Ilmiah," *Journal of Electrical and Computer Engineering*, 2015.
- [10] B. Herwijayanti, D. E. Ratnawati and L. Muflikhah, "Klasifikasi berita online dengan menggunakan pembobotan TF-IDF dan cosine similarity," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, pp. 306-312, 2018.

